

# **Deep sequencing of venom duct cDNAs from Indian cone snails.**

A Multi-institutional Project submitted for funding from the  
Department of Biotechnology

Project Coordinator: **K. S. Krishnan**  
Affiliation: **NCBS Bangalore**  
Address: **National Center for Biological  
Sciences TIFR GKVK Campus  
Bangalore 560065**

Participating Institutions and Investigators:

**National Center for Biological Sciences, Bangalore**  
K.S. Krishnan, R. Soudhamini

**Indian Institute of Science Bangalore**  
P Balaram, N. Balakrishnan

## **Indian Collaborators:**

**Center for Advanced Studies in Marine Biology Annamalai University**  
(Olivia Fernando)

**St Stephen's College Kollam** (Laladhas)

**Sugandhi Devadason Marine Research Institute, Tuticorin** (Murugan)

**Center for Fisheries Education, Mumbai** (Venkateshvaran)

**Andhra University Vishakapatnam** (Prabhakara Rao)

**Kaveri Medical Center Bangalore** (Kalyan Dewan)

## **Foreign Collaborators:**

Baldomero Olivera (University of Utah, USA)

Mani Ramaswami (Trinity College Dublin, Ireland)

**Agencies to which part of the project will be outsourced:**

**454 Life Sciences (Roche Group), Connecticut, USA. (see letter of intend)**

**Metaome Science Informatics (P) Ltd. Bangalore. (see letter of intend)**

## PROFORMA. I

### PROFORMA FOR SUBMISSION OF RESEARCH AND DEVELOPMENT PROJECTS, CREATION OF INFRASTRUCTURAL FACILITIES, CENTRES OF EXCELLENCE IN THE IDENTIFIED AREAS AND DEMONSTRATION PROJECTS

(To be filled by the applicant)

#### PART I: GENERAL INFORMATION

1. Name of the Institute/University/ Organization submitting the Project Proposal:

**National Center for Biological Sciences (Tata Institute of Fundamental Research)  
GKVK Campus, Bellary Road, Bangalore 560065.**

2. State: **Karnataka** 3. Status of the Institute: **Deemed University**

4. Name and designation of the Executive Authority of the Institute/University forwarding the application: **Prof. K. VijayRaghavan Director**

Project Title: **Deep sequencing of venom duct cDNAs from Indian cone snails.**

6. Category of the Project (Please tick): **R&D**

Demonstration

Establishment of Infrastructural facility/ Centre of Excellence

7. Specific Area (Please see Annexure - II): **Basic Research**

8. Duration: **Three Years**

9. Total Cost: **Rs. 96, 68, 000**

10. Is the project Single Institutional or Multiple-Institutional (S/M)? :

**Multi Institutional M**

11. If the project is multi-institutional, please furnish the following:

Name of Project Coordinator: **K. S. Krishnan**

Affiliation: **NCBS Bangalore**

Address: **National Center for Biological Sciences  
TIFR GKVK Campus Bangalore 560065**

12. **Project Summary** (Not to exceed one page. Please use separate sheet)

Cone snails are slow-moving, carnivorous marine molluscs that use a cocktail of about 100 venom peptides (conotoxins) to rapidly paralyze their often fast-moving prey. Conotoxins bind and affect function of key components of neurotransmission – including ion-channels and neurotransmitter receptors that are targets of existing and potential new neurotherapeutics. Thus, a calcium-channel targeting,  $\omega$ -conotoxin (under the tradename “Prialt”) is approved for treatment of chronic pain, and many others are in early or late stages of clinical testing.

Unusual evolutionary processes, implicit to predator-prey coevolution, have resulted in extremely rapid evolution of toxin sequences. Existing data indicate that each species contains an entirely unique complement of pharmacologically active conotoxins. These are estimated to fall into some 20 superfamilies (about half of which may have been identified) that share a distinctive leader sequences (pre- and pro peptides) and disulfide crosslinking patterns. However, current estimates, based almost entirely on biochemical and derivative molecular analyses, are yet to be confirmed by systematic DNA sequencing.

Using local bioinformatics support to organize and mine sequence information provided by a cutting-edge “454” high-throughput cDNA sequencing service, we propose to obtain the first deep sequencing of cDNAs expressed and/or enriched in venom ducts of a cone snail species, incidentally endemic to Indian coastal waters. While about 80 (of the roughly 500) *Conus* species are present in the Indian waters, we have selected *Conus araneosus* for four reasons. First, it is easily collected. Second, through extensive mass spectrometric, biochemical, structural and synthetic work conducted over the last 4 years, we have collected a large amount of proteomic information on this species. These pre-existing data can be used to evaluate the utility of this pilot deep sequencing project. In addition, we expect that the proposed work will allow us to correlate sequences of encoded peptides with pre-existing, but partially understood proteomic data. Third *C. araneosus* venom ducts are large enough to allow easy collection of starting polyA+ mRNA. And, finally, *C. araneosus* belongs to a poorly studied clade of the genus *Conus*, increasing the likelihood of discovery of novel conotoxin gene families.

This pilot project will: a) Provide massive scale cDNA sequence information to support an ongoing vigorous conotoxin research program; b) test the utility of this novel DNA sequencing technology for discovery of novel conotoxins or toxin modifying enzymes; and c) by establishing technical and intellectual infrastructure for this promising new approach, facilitate future use of this technology for gene discovery in any non-model organism of unusual potential biological or commercial interest.

### 13. PART II: PARTICULARS OF INVESTIGATORS

#### 1) Name: **P Balaram**

Date of Birth: **19/02/49**

Sex (M/F): **M**

Indicate whether Principal Investigator/Co-Investigator: **PI**

Designation: **Professor and Director**

Department: **Molecular Biophysics Unit**

Institute/University: **Indian Institute of Science**

Address: **IISc Campus Banaglore 560 012**

PIN: **560 012**

Telephone: **080-2932337, 080-3602741** Telex: **0845-8349 IISC IN**

Fax: **080-3600683, 0803600535**

e-mail: **pb@mbu.iisc.ernet.in**

No. of Projects being handled at present: 2 projects (DBT (Aquatic Biology), CSIR (NMITLI))

2) Name: **K. S. Krishnan**

Date of Birth: **19/06/1946**. Sex (M/F): **M**

Indicate whether Principal Investigator/Co-Investigator: **PI**.

Designation: **Professor**

Department: **Biological Sciences**

Institute/University: **National Center for Biological Sciences**

Address: **National Center for Biological Sciences TIFR GKVK Campus  
Bangalore 560065**

PIN: **560 065**.

Telephone: 080-23666123

Fax: 080 23636662

e-mail: ksk@ncbs.res.in

No. of Projects being handled at present: **Two (Peptides of therapeutic value from marine cone snails found in Indian Coasts, DBT supported Multi Institutional Project and Genetic studies of synaptic vesicle recycling TIFR supported)**

### **PART III: TECHNICAL DETAILS OF PROJECT**

(Under the following heads on separate sheets)

#### **16. Introduction**

Animal venoms used for predation and defense are highly evolved for potency and specificity. Biological processes that underlie the evolution of high-potency venom peptides have remarkable similarities to those used by medicinal chemists in the drug industry. Namely, to produce a lead compound, which then is randomly altered to generate "improved" higher potency variants, and which are selected for final use. Due to the billion years that evolution has had to generate and select high-performing variants, it has been argued that peptides/ molecules in wild species have been created with stringencies of selection rarely matched by medicinal chemistry.

Because venoms are also highly diverse, due to the evolutionary arms race, in which predators rapidly evolve new toxins to combat resistance mechanisms that evolve equally rapidly in prey, venom peptides represent an enormous untapped resource of biologically active compounds.

Of all venomous predators, the marine genus *Conus* has evolved the most amazing repertoire of toxins. Each of the more than 500 species of cone snails thought to exist produces at least 100 if not 200 distinct, highly specific, neuroactive peptides. The high specificity of these peptides for specific channel isoforms makes them invaluable and widely used tools for neurophysiology. More spectacular is the commercial potential of these unique compounds. At least 6 of them are currently in clinical trials for pain, epilepsy or cardiac disorders. But one, a component of *Conus magus* venom, is used under the tradename "Prialt" to manage chronic pain that cannot be treated by morphine.

Each of these peptides unlike many other animal peptides is coded by a single gene. Members of each family of the *Conus* peptides are coded by genes with highly conserved pro- and pre-peptide sequences which are identical for each gene family expressed within a species, and highly conserved across species. In contrast, venom peptide sequences which show a strikingly increased rate of mutation, perhaps five times as fast as the rate of conventional or silent mutations. This apparently high mutation rate, which may represent either positive “diversifying” selection for altered peptides, or a biological process of hypermutation as observed for immunoglobulin genes, results in a high diversity of conopeptides within and across *Conus* species.

An additional feature of cone snail peptides is that they perhaps are the most post translationally modified with almost as many as six modifications found in a ten amino acid peptide. Modifications like Bromo tryptophan, very common to conopeptides, as well as sequence specific epimerization to create D-aminoacids, have been thought to be unique to conotoxins (although recent reports suggests that Tryptophans can be brominated in some mammalian cells). Although the enzymes that mediate these post-translational modifications are to be of considerable basic and applied interest (with the potential to expand biological activities of commercially produced proteins), they so far remain largely uncharacterized at a biochemical or molecular level.

Given the biology of conopeptides outlined above, it is easy to appreciate that the characterization and synthesis of genes highly expressed in venom ducts of *Conus* may yield a “pharmacopia” of over 100,000 highly optimized compounds of considerable biological and applied interest. The biggest payoff from such an effort will be the identification of lead compounds to treat pain, epilepsy and heart disease (to name a few: note the unexpected uses for “Botox”). However, there will also be many, concurrent and more guaranteed payoffs detailed later in this section.

Despite the growing conviction that bioactive peptides in general, and conotoxins in particular, are a hugely interesting and untapped natural resource, there has not been a strong committed attempt at massive scale genome sequencing to identify conotoxin and conotoxin production genes. There are two reasons for this lacuna in the field. First, is the need for an interdisciplinary, human infrastructure in which expertise in DNA sequencing, bioinformatics and biology are brought together to focus on this problem. This has not been easy for the 3 or 4 other major conotoxin groups around the world. Second, has been the high cost of such a project although, at present, we argue costs are no longer as prohibitive as they have been in the recent past.

The Indian conotoxin group is uniquely positioned to take on this challenge for several reasons. First, in addition to natural resources (80 *Conus* species in India) we have a naturally interdisciplinary and cohesive group of scientists who represent expertise in Marine Biology, Molecular Genetics, Peptide Chemistry, Structural Biology, Evolutionary Biology, Neurophysiology, and Neuroscience. We add to this, local expertise in Bioinformatics. Second, we are ready to approach this project at a time when the technology for massive cDNA sequencing may have just reached the level required for this effort to be successful. Using modern array-based pyrosequencing technologies, a sequencing service will provide us with about 700,000 sequencing reads – each 400 bases long, for about Rs 26 lakhs. Based on numbers obtained using earlier, less developed versions of this technology, we estimate that the 454

Assembler software (Newbler) should provide us with “contigs” for about 3000 cDNAs, each constructed from about 20 independent sequence reads.

We have chosen to analyze cDNA (rather than genome) sequence for two reasons. First, we are particularly interested in genes expressed at high levels in the venom duct. Second, because a typical mollusc genome is more than twice the size of the human genome, even current technologies for genome sequencing will incur considerable costs – currently about Rs.15 crores.

While, other technical issues are considered later in this proposal, it is important to introduce at this point, our choice of the first *Conus* species to analyze. Of the 80 or so *Conus* species reported in the Indian waters, our selection is based on several criteria. A. The ease of collection and availability of the species. B. The size of the venom ducts, larger sizes guaranteeing sufficient yield of polyA<sup>+</sup> mRNA; C. Evolutionary distance from more intensively studied Pacific species; greater distance increasing the likelihood of unique discoveries; and D. Existing proteomic/ mass spectrometry data accumulated by the Indian conotoxin group, which would allow immediate use of sequence information when obtained.

Of the three species that represent a good balance of these criteria (*C. monile*, *C. amadis*, and *C. araneosus*) we have selected *C. araneosus* for first analysis.

A massive scale analysis of cDNA sequences expressed in the *C. araneosus* genome has the potential to achieve the following goals.

- a) Allow identification of genes that encode new conotoxins and potentially new families of conotoxins
- b) Allow mechanisms (enzymes) involved in conotoxin production to be identified and subsequently harnessed for heterologous protein production.
- c) Expose/train young Indian biologists and chemists to important and multidisciplinary science that uses emerging technologies.
- d) Test in the Indian scientific context, and develop the infrastructure to exploit, important new technologies for high throughput DNA sequencing.

## 16.1 Origin of the proposal :

A multi institutional collaboration to isolate and characterize *Conus* peptides was begun in 2001 by Profs K S Krishnan and P Balaram, supported largely by institutional funds. Initiated on a relatively small scale, we soon realized the potential of this discovery and applied for and received a modest DBT grant award, which now constitutes the main funding for the ongoing work. This work allowed us to put in place people and processes required for *Conus* collection and well as biochemical, structural, synthetic and proteomic analyses of conopeptides. In particular, we have: a) identified locations and seasons for collecting about 40 of the 80 described Indian cone snail species; b) used HPLC and MALDI mass-spectrometry to profile peptides present in the venom

ducts of several species, in particular those from *C. araneosus*, *C. monile*, and *C. amadis*; c) characterized post-translational modifications on a subset of these peptides, some of the enzymes involved, and synthesized biologically active forms of an even smaller subset of these conopeptides.

We have also, with help from Uma Ramakrishnan in the NCBS, studied the evolution of *Conus* and mapped different *Conus* species into clades based on the relatedness of their cytochrome C sequences. Together with the involvement of neurophysiologists (S.K. Sikdhar and M.K. Mathew) and neuroscientists (S. Chatterjee), the project has developed to a stage at which we are capable of using and exploiting systematic knowledge of conotoxin sequences. Thanks to the advancement in array base pyrosequencing technologies, we believe that this is the appropriate time to take on a large-scale cDNA sequencing project.

## 16.2 Definition of the problem

1. To sequence and identify cDNAs expressed and enriched in the venom duct of cone snails.
2. To use this sequence to guide peptide identification, synthesis and characterization.

## 16.3 Objectives:

The major objective of this project is to analyze and characterize large-scale cDNA sequence information from the venom duct of a cone snail endemic to India (*C. araneosus*). In doing so, we will enhance indigenous capabilities in sequence assembly, annotation and analysis and provide challenging opportunities to scientific trainees and young scientists. The work also has the potential to greatly energize molecular marine biotechnology in particular and biodiversity-based molecular discovery programs in general.

# 17. Review of Current Status of research and development in the subject

## 17.1 International Status:

Despite obvious interest in *Conus* venoms, pioneered by Baldomero Olivera, a direct cDNA sequencing approach to gene discovery has never been performed. Pilot cDNA sequencing by Olivera and colleagues were not as useful as expected, largely because they were initiated when the average read length for pyrosequencing was about 90 base pairs (Olivera, personal communication). Very recently, the European Union funded a consortium grant application for a multidisciplinary *Conus* program that included a proposal to sequence the genome of *Conus quercinus*. However, in part due to the death of their lead scientist and in part due to the somewhat non-cohesive nature of the consortium, this project has not yet gathered significant momentum. Molluscan genomes have hitherto been ignored by large sequencing projects. However, the *Aplysia californica* (sea slug) and *Biomphalaria sabralta* genomes are

being sequenced. A 2X draft assembly of the *Aplysia* genome sequence is available on the NCBI database and it is expected to be completely annotated and searchable in a few months. This will be a useful (though non-essential) resource for annotation of *Conus* cDNA sequences when obtained.

## 17.2 National Status:

Other than stray reports of *conus* availability and a study or two of crude venoms and radula teeth all serious study of *conus* peptides, sequence information and proteomics has come from the group coordinated by Profs. Balaram and Krishnan, who are the leading members of this consortium for deep sequencing.

## 17.3 Importance of the proposed project in the context of current status:

Cone snails belong to a large genus of molluscs of considerable biological and applied interest. Our Indian consortium of *Conus* scientists is uniquely poised, not only due to the considerable natural resources in Indian waters, but also due to our blend of biological skills to make important new discoveries. In addition, due to the paucity of focussed DNA sequence and bioinformatic analyses conducted in India, this project will represent a good test case for efficient large scale DNA sequencing and data mining in India.

## 17.4 Anticipated Products & Processes of Practical/Technological utility/Socio-economic relevance expected to be evolved by pursuing the project.

These have been partially outlined at the end of the Introductory section.

a). The most optimistic outcome would be the discovery of lead drugs which therapeutic effects on rodent models for pain, epilepsy, or cardiac disorders.

However, more guaranteed products/processes of value in the Indian context are:

b) Sequences of new conotoxin genes and, potentially, new conotoxin gene families

c) Sequences of genes encoding conotoxin production enzymes, including those that mediate protein processing, folding and post-translational modification. These have the potential to be useful within and outside the field of conopeptide research.

d) Exposure and training of young Indian biologists and chemists in important and multidisciplinary science that uses emerging technologies. Indeed, one of the most pleasing features of the conotoxin project so far has been the new intellectual vistas it has opened to chemists exposed to biodiversity based natural products, and to marine biologists, exposed to the power of chemical and molecular technology.

e) Establishment of a local infrastructure to use important new technologies for high throughput DNA sequencing for gene discovery.

This project using cutting edge array based pyrosequencing will be a pilot program to test the value of this highly anticipated approach to biodiversity based gene discovery.

#### 17.5 Expertise available with the proposed investigating group/institution in the subject of the project.

This group has proven expertise in conotoxin research needed for making progress in this project. In particular, methods for *Conus* specimen collection, duct isolation, RNA extraction and biochemical, structural, proteomic and neurophysiological analyses are well established as demonstrated by the publications listed below.

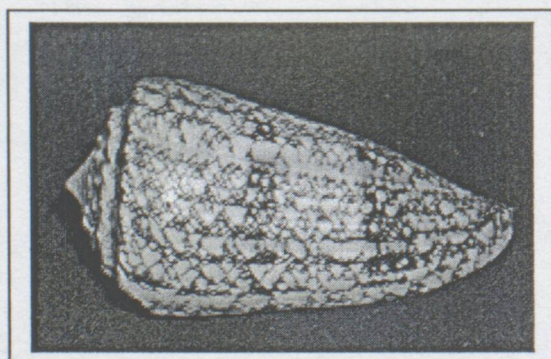
The NCBS and the PI's have considerable working expertise in genome biology. DNA sequencing as well as contig assembly will be conducted by 454 Life Sciences, the world leader in array-based parallel pyrosequencing. Subsequent refined contig assembly (if required) as well as DNA sequence organization, annotation, and analysis will be conducted by a Bangalore Bioinformatics company, Metaome, a recent recipient of a DBT grant, whose founder Ramkumar Nandakumar (CV attached) has considerable experience in these procedures based on his previous experience in a Max (Tuebingen) and in the Gurdon Institute (Cambridge).

#### 17.6 List of 5 experts in India in the proposed subject area:

SN	Name	Designation	Address
1.	S E Hasnain	VC	Hyderabad Central University
2.	V S Chauhan	Director	ICGEB, Delhi
3.	S. Brahmachari	DG	CSIR, Delhi
4.	Siddhartha Roy	Director	IICB, Kolkata
5.	J. Gowrishankar	Director	CDFD, Hyderabad
6.	Girish Sahni	Director	IMTECH, Chandigarh

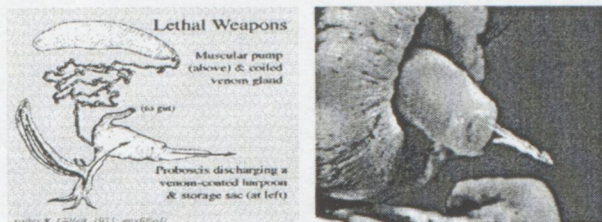
### 18. Work Plan and Methodology:

#### 18.1. *C. araneosus* collection and venom duct storage.



We will collect *C. araneosus* from the Tamil Nadu coast and dissect out their venom ducts on site using standard methods. The ducts will be stored in a *RNAlater* (Ambion), a medium that keeps the RNA stable and in good condition for weeks. If this compromises

Venom apparatus of a Cone



Courtesy: Australian Marine Shells, v.1, v.2 Odyssey Publishing, San Diego, CA, 377 pp, 97 color plates, c. 1993, c. 1994.

the quality of cDNA required (see next section), then we will transport the snails in a salt water tank to NCBS, where they will be dissected and tissue extracted for immediate RNA isolation and cDNA synthesis. Specimen collection will be conducted with help with our long-time collaborators, Olivia and Anthony Fernando at the Center for Advanced Studies in Marine Biology, of Annamalai University.

### 18.2. cDNA synthesis from Venom duct.

The venom duct is a long tube connected to a bulb which acts as a muscular pump that propels venom in the duct into the prey via a terminal "harpoon." The duct epithelium contains cells that express genes for conotoxin and all enzymes and processes required for their production, maturation and secretion into the duct lumen. We estimate, but need to ascertain, that about 2 gms of venom duct tissue will be enough to yield the 5-10 micrograms of polyA + mRNA ideally required for production of the 5 micrograms of single stranded cDNA that will serve as the starting material for pyrosequencing.

RNA will be isolated using Ambion's "ToTALLY RNA" isolation kit, optimized for 0.5-10gms of tissue, followed by polyA RNA purification using Ambion's "Poly(A) Purist" kit. We will check yields and quality of mRNAs and then use them for cDNA synthesis. For relatively large-scale synthesis of first strand cDNA we will use either the Superscript III system (Invitrogen) and/or the Monsterscript system (Epicentre Biotechnologies) with oligo-dT priming as this is likely to enhance the frequency at which we obtain sequences of conotoxins. All of the above procedures and subsequent Qiagen column purification will be optimized for high yields of pure cDNA as per specifications from 454 Life Sciences (5micrograms of cDNA, at 300ng/ml in TE at purity of  $A_{260}/A_{280}$  of ~1.8). If RNA is limiting for any reason, then there are appropriate ways to amplify polyA+ RNA under conditions that allow their initial representation to be maintained (Clontech SMART RNA amplification kit).

Ultimately, we propose to provide 454 Life Sciences with two cDNA samples.

First, cDNA simply created by oligo-dT priming of purified polyA+ RNA. An advantage of this approach is that in addition to sequence data, analysis of such a cDNA sample should give us information on the relative expression levels of each mRNA (by counting the relative frequency at which each EST is encountered). A disadvantage is that high and moderately expressed RNAs will predominate, and RNAs expressed at lower levels will be relatively sparsely sampled. The second cDNA sample, will address this limitation.

Second, we will provide a sample of "normalized" cDNA in which by controlled RNA / cDNA hybridization, highly expressed RNAs will be subtracted out using kits specifically designed for such normalization (eg, Trimmer / Trimmer Direct from Evrogen). 454

sequence of this normalized cDNA sample will allow us to obtain conotoxin sequences expressed at lower levels, which is important given the potentially high payoff of identifying novel conotoxin families or superfamilies.

**18.3. 454 array based parallel pyrosequencing.** The pyrosequencing technology ("sequencing by synthesis") was invented in the early 1990s and eventually licenced to 454 Life Sciences, now a subsidiary of Roche. At 454, the technology was adapted for use in nano-arrays to allow upto a million parallel pyrosequencing reactions to be performed withi 7 hours. For massive cDNA sequencing, cDNA derived from tissue is fragmented in gaseous phase, and coupled via ligated primers to nanobeads under conditions where each bead has either one or no attached DNA molecule. The beads are dispensed into nanotitre dishes, each well containing one bead. Individual cDNAs in each well are amplified using PCR. Thus, before array pyrosequencing, each nanowell contains multiple copies of a unique template DNA attached to a polymer bead.

The basis for the technique and current results from its use are hugely impressive and overviewed in the websites below. The sites also include a large number of references that use the 454 sequencing services.

Overview: <http://www.454.com/>

Publications: <http://www.454.com/news-events/publications.asp>

cDNA sequencing and whole genome survey sequencing publications:

<http://www.454.com/news-events/publications.asp?cat=14>

<http://www.454.com/news-events/publications.asp?cat=4>

Application notes: <http://www.454.com/sequencing-services/protocols.asp>

It is important to note that many of the publications above are based on the earlier generation machines and reaction conditions with 90 to 130 bp reads, which make contig assembly substantially more difficult.

With technology we intend to use in this project, when one factors in a failure rate for wells that contain beads without template DNA, the currently anticipated 400 bases per sequence read allows about 250 MB of sequence to be obtained in two parallel sequence runs, completed within 7 hours of operation of a sequencing unit

We propose to use two runs because of anecdotal information (based on prior experience at 454 Life Sciences) which indicates that for cDNA/ EST sequence this represents a threshold beyond which one obtains diminishing returns in terms of independent sequence information that helps in definining/ refining genes/ contigs. However, the diffrence information from a single run (125MB) and two runs is substantial.

The current cost for creation of the bead-associated cDNA library, two sequencing runs, contig assembly and software access is €37,700 (See letter from the company).

#### **18.4. Assembly and annotation phases.**

454 Life Sciences will provide us with : i) unordered set of sequences of assembled contigs in fast A format with associated quality scores; ii) all sequence reads in fast-A format with an associated quality scores. iii) A licence for their in house DNA assembly, mapping and data analysis software.

Beyond this, Dr. Ramkumar Nandakumar, from Metaome Inc, Banaglore, will supply and be responsible for Bioinformatics support for this project. Metaome will organize a database on Sun workstations, that will also run several key Bioinformatics programs locally. Thus, they will install a suitable genome annotation platform on the central server, organize databases with sequences from 454, as well as downloaded genome sequences from the NCBI and/or other genome databases. Their first main task will be to use local serial blast analyses to annotate transcripts (~5000) by serial blast analyses. Second phase will consist of more detailed analysis.

These will include careful analysis of levels of transcription of genes, providing easy output of sequence quality in the contigs and/or ESTs. In addition, they will provide custom sequence alignment and motif search analyses to discover and/or analyze conotoxin/ processing and modification enzymes. They will provide custom software with convenient interfaces that generates outputs of interest to the specific downstream conotoxin researchers. For example, they may also be used for designing PCR primers one may wish to use to amplify and clone full-length clones and or new family/superfamily members from identified conotoxins. They may virtually fold novel conotoxin sequences around template scaffolds identified by structural analysis of homologous, previously studied conotoxins.

In addition, and most important immediately after the sequence assembly, they will provide customized outputs to be used for more detailed annotation. For this detailed annotation phase, we propose invite at least 4 international experts on conotoxins and/or post-translational modification enzymes, whose expertise will be harnessed with immediate bioinformatics support.

Expectations: a) We expect to identify DNA sequences encoding most of the conotoxins expressed in the venom duct, even if these are not full length. This is justified by the argument that with reads of 400bp, from a oligo-dT primed cDNA library, 3' cDNA sequences encoding short peptides certainly be represented. b) ) We expect to obtain about 2000 different full-length contigs for genex expressed in the venom duct as well as information on their relative abundance. The contigs will be assembled on the basis of sequence overlap (Newbler) as well as parallel BLAST alignments with homologous sequence in other species eg. *Aplysia*. The number of contigs is estimated on the basis of several arguments most effectively summarized by considering a similar analysis EST sequences from a non-model organism, the Glanville fritillary butterfly [Vera, Wheat et al., Molecular Ecology, 2008].

Vera et al. used early 454 technology to obtain and analyze 600,000 reads of 110 bases each of ESTs from the butterfly. Despite an average contig length of less than 300 base pairs and a very low frequency of large contigs obtained, the authors were: a) able to match a very large fraction of these to homologs in other insects; and b) estimate that if they had 4-fold increase in data (number of 110bp ESTs) , they would

have been able to create full-length contigs for 50% of the genes that had homologs in other species.

Our analysis should hugely improve on this (quite respectable) study for four reasons:

1) Due to the increased length of each sequence read, our unit EST size (400bases) is 4 times bigger than used in the above study.

2) Given that we will have a larger number (1,400,000 versus 600,000) of total reads, we should have a total of 8-times as much sequence information. This should allow more efficient assembly of contigs, which are composed of larger sized units.

3) Finally, because the mature conotoxin sequence is encoded in the 3' end of mRNA coding sequence, these most important sequences should be generally contained within a single sequence read from the 3' end of the mRNA. Indeed, this is the reason why we choose to create initial cDNA using oligodT priming, rather than random oligonucleotide priming.

4. Initial sequence assembly will be carried out by 454 Life Sciences using their "Newbler" assembler software designed with knowledge of the specific sequence errors and ambiguities implicit to the pyrosequencing technique. In contrast, Vera et al. used publicly available Assembler software, which, though efficient, may not have corrected specific subtypes of sequence errors.

#### **18.5. Sequence correlation with Mass spectrometry.**

The presence of peptides predicted by cDNA analysis will be established by direct MALDI-MS/MS de novo sequencing of peptides in the crude venom. The determination of partial peptide sequences and comparison with predicted sequence will also help in characterisation of post translational modifications and sites of hydroxylation etc. In addition predicted peptides will be synthesized and compared to crude venom peptides by both HPLC and MALDI analysis to establish sites of amidation, isomerisation etc. These are also part of an ongoing DBT project on analysis of venom peptides from *Toxoglossa* species found in Indian waters.

#### **18.6 Outcome expected from the project.**

NCBS and IISc will hold patents for peptides and sequences we find and characterize biologically. Licence will be given to Pharma industry to pursue trials and for exclusive or nonexclusive use

#### **18.7 Time schedule of activities giving milestones:**

We expect, in a short time after the project is approved, starting to get sequence data and hope deep sequence data on venom glands from at least two species of snails will be obtained in the first year. The second year will be mostly spent in analysis and organising data and in addition we may start sequencing yet another snail species or go for whole genome sequencing or sequencing cDNA from unrelated tissues from the same species in an attempt to get at enzymes involved in post translational modification of the expressed peptides. The second and third years also will witness a

large scale attempt at identifying new families of peptides, isolation and synthesis of new peptides etc.

### 18.8 Project implementing Agency/Agencies

**Name of Agencies:** National Center for Biological Sciences and Indian Institute of Science

**Address of Agencies:** NCBS, Tata Institute of Fundamental Research, GKVK Campus, Bangalore 560065  
And IISc, Bangalore 560 012

**Proposed Budget:** Rs. 96,68,000

**Large scale cDNA sequencing (454 Life Sciences):**  
Rs. 53.6 lakhs (€80,000)

**Bioinformatics Support (Metaome):**  
Rs. 3.9 Lakhs

**Server for storing sequence files and analysis programs:**  
Rs. 2.0 Lakhs.

**Travel:** Rs. 3.7 Lakhs  
(Two visits each for collaborators from US and Ireland); 6 local round trip flights for Pls.

**Personnel:** One project Scientist, One postdoc and one JRF – each for 3 years.  
**Salary component:** Rs. 19.08 lakhs

**Consumables and Contingency**  
12.00 lakhs

**Overhead:** Rs. 2.40 lakhs

**Total** Rs. 96.68 lakhs

## BUDGET

### PART IV: BUDGET PARTICULARS

Grand Total Requirement for all institutions put together :)

#### I. Budget for IISc ( P Balaram,)

##### A. Non-Recurring (e.g. equipments, accessories etc.) NIL

##### B. Recurring

##### B.1 Manpower

Project Scientist @ 25,000 PM three years: 9, 00,000

Post Doctoral: @ 16,000 PM three years: 5, 76,000

JRF : @ 12,000 PM three years: 4, 32,000

Total: 19, 08000

B.2 Consumables 3 X 3 lakhs= 9 Lakhs

Other Items	Year 1	Year 2	Year 3	Total
B.3 Travel				
B.4 Contingency 3X 1 lkhs= 3 Lakhs				
B.5 Overhead Charges (15%) 1. 85 lakhs				
Sub-Total (B= B.3 + B.4 + B.5) 4.35 lakhs				

Sub-Total (B= B.1+ B.2 +B.3 + B.4) = 32, 93,000

Grand Total (A + B) = 32, 93,000

#### II Budget for NCBS/DBS TIFR (K S Krishnan,)

##### A. Non-Recurring (e.g. equipments, accessories etc.)

S.No
1. 454 Life Sciences payment Rs 53.6

lakhs
2. metaom e payment Rs 3.9 lakhs
3. Server 2.00 lakhs

**B. Recurring****B.1 Manpower**

<b>Sr. No.</b>
1.
<b>Sub- Total (B.1)</b>

**B.2 Consumables**

Sr. No.	Item	Year 1	Year 2	Year 3	Total
1					
	<b>Sub- Total (B.2)</b>				

Other Items	Year 1	Year 2	Year 3	Total
B.3 Travel 3. 7 lakhs				
B.4 Contingency				
B.5 Overhead Charges (15%) 0.55				
<b>Sub-Total (B= B.3 + B.4 + B.5) 4.25</b>				

Sub-Total (B= B.1+ B.2 +B.3 + B.4) = 4. 25 lakhs  
Grand Total (A + B) = 63. 75 lakhs

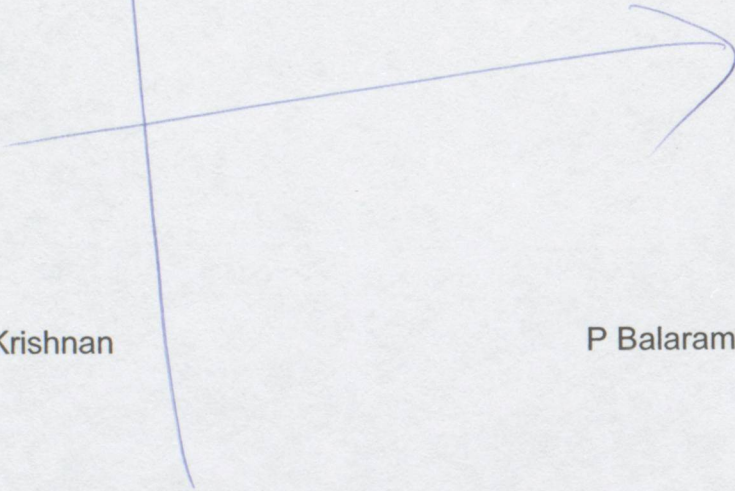
Place: Bangalore

Date: June 17 th 2008

Signature of Investigator(s)

K. S. Krishnan

P Balaram



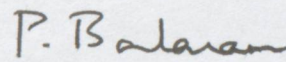
Place: Bangalore

Date: June 27<sup>th</sup> 2008

Signature of the Investigator(s)



K. S. Krishnan



P. Balaram

### Jusitification:

The major cost is payment to the two companies which will do the sequencing and analysis. These companies have great expertise in sequencing and analysis and it is both cost effective and time effective to outsource these routine procedures to them. Although routine in their hands it will be an enormous task organising to do sequencing in our own setups and will be wasteful in terms of time and talent. On the oher hand it is best we focus our efforts on the next phase after the sequences have been obtained of annotating and identifying useful peptides. These will be primarily done by the PI's and the project scientist with the help of the post-doctoral fellow and the JRF. The project scientist and post doc will be involved in the annotation and follow up. The JRF will be needed for sample collection and cDNA preparations in addition to helping the project scientist in annotation. Travel is for the collaborators both foreign and Indian to visit Bangalore and for PI's to travel in side the country for the annotation jumborie which is planned.

### International Collaborators:

We are fortunate that the very founder of cone snail venom research Prof. Baldomeiro Olivera has agreed to be a collaborator on this project. Although in his inmicable and humble style he states he has little to offer, from our personal interactions we know he is a vast fund of information on venoms and will go along way in making this effort a success. Toto in addition has been very considerate in helping us evolve our strategies on conotoxin research and has been a well wisher and admirer of our two institutions. We hope through this collaboration he will be able to assist in the speedier annotation that is needed to beat the competition from Europe.

Dr. Ramaswami has had long standing collaborations with the project coordinator in Drosophila Neurobiology. He was in fact the one who propelled the conotoxin project in the first instance by dragging some of us to the sea coast and financing from personal funds our early collection trips. He has continued to evince a keen interest in the conotoxin project. He intiated the discussion on this project and obtained all the details on the sequencing strategies. Being an adjunct faculty at TIFR he will be very much a part of our effort in future as well.

## PART V: EXISTING FACILITIES

20. Available equipment and accessories to be utilized for the project:

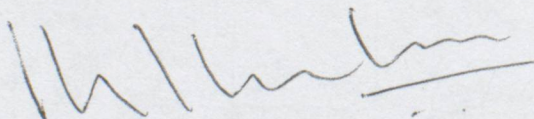
SN	Name of equipment/ Accessories	Name	Model	Funding Agency	Year of Procurement
1	Patch Clamp set up	Axoclamp with O clamp software		DST	2000
2	HPLC	Various			
3	MALDI GC/MS				
4	NMR				
5	Various other				

## PART VI: DECLARATION/CERTIFICATION

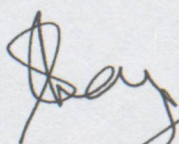
It is certified that

- (a) The research work proposed in the scheme/project does not in any way duplicate the work already done or being carried out elsewhere on the subject.
- (b) The same project has not been submitted to any other agency/agencies for financial support.
- (c) The emoluments for the manpower proposed are those admissible to persons of corresponding status employed in the institute/university or as per the Ministry of Science & Technology guidelines (Annexure-III).
- (d) Necessary provision for the scheme/project will be made in the Institute/University/State budget in anticipation of the sanction of the scheme/project.
- (e) If the project involves the utilisation of genetically engineered organism, it is agreed that we will ensure that an application will be submitted through our Institutional Biosafety Committee and we will declare that while conducting experiments, the Biosafety Guidelines of the Department of Biotechnology would be followed in toto.
- (f) If the project involves field trials/experiments/exchange of specimens, etc. we will ensure that ethical clearances would be taken from concerned ethical Committees/Competent authorities and the same would be conveyed to the Department of Biotechnology before implementing the project.
- (g) It is agreed that any research outcome or intellectual property right(s) on the invention(s) arising out of the project shall be taken in accordance with the instructions issued with the approval of the Ministry of Finance, Department of Expenditure, as contained in Annexure-V.
- (h) We agree to accept the terms and conditions as enclosed in Annexure-IV. The same is signed and enclosed.
- (i) The institute/university agrees that the equipment, other basic facilities and such other administrative facilities as per terms and conditions of the grant will be extended to investigator(s) throughout the duration of the project.
- (j) The Institute assumes to undertake the financial and other management responsibilities of the project.

Signature of Project Coordinator  
(Applicable only for multi-institutional projects)



Signature of Executive Authority of Institute/University with seal



Date :

1 July 2008

**K. VijayRaghavan**  
Director  
National Centre for Biological Sciences  
Tata Institute of Fundamental Research  
GKVK, Bellary Road, Bangalore-560 065